

## Data Collection

The data required for this project is as follows has been collected from multiple sources. The following data is required for the project:

- 1) Neighborhood data of Mumbai
- 2) Geographical coordinates of Mumbai and all neighborhoods in Mumbai
- 3) Venue data for neighborhoods in Mumbai

## Neighborhoods Data

The data of the neighborhoods in Mumbai was scraped from [https://en.wikipedia.org/wiki/List\\_of\\_neighbourhoods\\_in\\_Mumbai](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai). The data is read into a pandas data frame using the `read_html()` method. The main reason for doing so is that the Wikipedia page provides a comprehensive and detailed table of the data which can easily be scraped using the `read_html()` method of pandas.

## Geographical Coordinates

The geographical coordinates for Mumbai data has been obtained from the GeoPy library in python. This data is relevant for plotting the map of Mumbai using the Folium library in python. The geocoder library in python has been used to obtain latitude and longitude data for various neighborhoods in Mumbai. The coordinates of all neighborhoods in Mumbai are used to check the accuracy of coordinates given on Wikipedia and replace them in our data frame if the absolute difference is more than 0.001. These coordinates are then further used for plotting using the Folium library in python.

## Venue Data

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighborhoods in Mumbai and is used to study the popular venues of different neighborhoods.